

Opinnäytetyö (AMK)

Tietojenkäsittely

Yrityksen tietojärjestelmät

2014

Eveliina Gratscheff

MYYNNINENNUSTEIDEN ANALYSOINTI SAP:N TIEDONLOUHINTAMALLEILLA



TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

Eveliina Gratscheff

MYNNINENNUSTEIDEN ANALYSOINTI SAP:N TIEDONLOUHINTAMALLEILLA

Opinnäytetyön tavoitteena on selvittää onko SAP-järjestelmän tiedonlouhintatyökaluja mahdollista hyödyntää mynninennusteiden tarkkuuden analysoinnissa. Työssä on tarkoitus joko kehittää prosessi ja malli, joiden avulla mynninennusteiden tarkkuutta voidaan mitata, tai päätellä, ettei tiedonlouhintaa voi käyttää tämän tyyppiseen analyysiin.

Teoriaosuus koostuu tiedonlouhinnan teoriasta keskittyen erityisesti tiedonlouhintaan SAP-järjestelmässä. Käytännön osuuden tavoitteiden tueksi on kuvattu mynninsuunnittelu SAP-järjestelmässä. Käytännön osuus suoritetaan tapaustutkimuksena perustuen toimeksiantajan tarjoamaan testidataan.

Tutkimuksen lopputulos on se, että ABC-luokitusta ja painotettuja pisteytystaulukoita voi käyttää mynninsuunnittelun onnistumisen arviointiin. Jos analyysi voidaan suorittaa yhden tunnusluvun, esim. toteuman ja ennusteen välinen ero, perusteella, paras vaihtoehto on ABC-luokitus, joka luokittelee objektit käyttäjän määrittämiin luokkiin. Jos on tarpeen käyttää useampia tunnuslukuja tai ominaisuuksia niiden lisäksi, voidaan analyysi suorittaa painotetulla pisteytystaulukolla.

Opinnäytteessä on kuvattu ratkaisu, jonka avulla haluttuja asioita on mahdollista analysoida. Lisäksi on perusteltu, miksi muut tiedonlouhintamallit eivät sovellu halutun asian analysointiin.

ASIASANAT:

Tiedonlouhinta, mynninennusteet, SAP.

Eveliina Gratscheff

ANALYZING DEMAND PLANS WITH SAP DATA MINING MODELS

The objective of the present bachelor's thesis is to determine if the data mining models provided by SAP ERP can be used to analyze demand plans. The aim of the study is to either find a model and a process which can be used in defining the accuracy of the plans, or to conclude that data mining cannot be used for this kind of purpose.

The theory part of the thesis consists of a description of data mining with a special focus on data mining in SAP. In addition, to support the objective of the empirical part, the demand planning process is explained. The empirical part is conducted as a case study based on an example dataset provided by the client company.

The conclusion is that the ABC classification and weighted score tables can be used to analyze the success of demand planning. If the analysis can be based on one key figure, for example the difference between the forecast and the history, the ABC classification is the best alternative. If it is necessary to use several key figures or key figures and characteristics, weighted score tables can be used to take all of them into consideration.

The present thesis describes a solution that can be used to analyze the accuracy of demand planning. In addition, the study explains why other models cannot be applied to this type of analysis.

KEYWORDS:

Data mining, demand planning, SAP.

SISÄLTÖ

SISÄLTÖ	4
1 JOHDANTO	6
2 TIEDONLOUHINTA	8
3 TIEDONLOUHINTAPROSESSI	10
4 TIEDONLOUHINTAMETODIT	17
4.1 ABC-luokitus	18
4.2 Pisteytys	18
4.3 Painotettu pisteytystaulukko	19
4.4 Yhteenkuuluvuusanalyysi	20
4.5 Päästöpuu	21
4.6 Ryvästys	22
5 MYYNNINSUUNNITTELU SAP APO DP:SSÄ	23
6 ENNUSTEVERSIONOIDEN ANALYSOINTI	26
6.1 Tehtävän määrittely	26
6.2 Valmistelut	26
6.3 Datan käsittely	29
6.4 Mallin valinta	31
6.5 ABC-luokituksen tutkiminen	32
6.6 Testaus	38
6.7 Valmiit prosessit	40
7 YHTEENVETO	43
LÄHTEET	45

KUVAT

Kuva 1. Esimerkki päätöspuusta.....	21
Kuva 2. Versiodataa listattuna sellaisena kuin se näkyy valmiissa infokuutiassa.	28
Kuva 3. Toteumadataa listattuna sellaisena kuin se on valmiissa infokuutiassa.....	29

Kuva 4. Analyysiprosessi tuotekohtaiselle painotetulle keskiarvolle.	40
Kuva 5. Analyysiprosessi asiakaskohtaiselle painotetulle keskiarvolle.	41
Kuva 6. Tuotekohtaisen painotetun keskiarvon raportti.	42

KUVIOT

Kuvio 1. Myynninsuunnitteluprosessi (SAP Help Portal 2013h).....	24
--	----

TAULUKOT

Taulukko 1. Tuotekohtaisen ennusteen tarkkuuden luokat.....	35
Taulukko 2. Painotetulla keskiarvolla lasketun tuotekohtaisen ennusteen tarkkuuden luokat.....	36
Taulukko 3. Painotetulla keskiarvolla lasketun asiakaskohtaisen ennusteen luokat. ...	37
Taulukko 4. Testausvaiheessa painotetulla keskiarvolla lasketun tuotekohtaisen ennusteen tarkkuuden luokat.	38
Taulukko 5. Testausvaiheessa painotetulla keskiarvolla lasketun asiakaskohtaisen ennusteen luokat.	39

1 JOHDANTO

Opinnäytetyön aiheena on tutkia soveltuvatko SAP-järjestelmän tarjoamat tiedonloughinnan työkalut myynninsuunnittelussa luotujen ennusteiden onnistumisen arviointiin. Työn tavoitteena on joko löytää ratkaisu, jolla versiodataa voi analysoida tai sulkea pois tiedonloughinnan käyttö tällaisessa analyysissä. Suunnitelmaversiot ovat peräisin Advanced Planner and Optimizer -järjestelmän myynninsuunnittelusta ja ne sisältävät tuotteiden myynnin ennusteita. Opinnäytetyössä pitäisi selvittää onko mahdollista mitata ennusteen tarkkuutta, jonkin versiodatassa olevan ominaisuuden perusteella. Sen lisäksi on testattava eri metodien toimintaa ja mallien kofigurointiin käytettäviä asetuksia.

Työn teoreettinen viitekehys koostuu tiedonloughinnanteoriasta erityisesti SAP-järjestelmän näkökulmasta täydennettynä tiedonloughinnan teorialla yleisellä tasolla. Lisäksi käytännönsuuden tavoitteiden tukemiseksi kuvataan myynninsuunnittelua SAP APO DP:ssä. Tiedonloughinnassa on keskitytty erityisesti prosessin kulkuun ja käytettävissä olevien mallien kuvaukseen. Teoriaosuus perustuu alan kirjallisuuteen sekä Internetlähteisiin.

Tutkimukseen on valittu SAP-järjestelmän tarjoamat tiedonloughintametodit. Ulkopuoliset tiedonloughintamenetelmät, joita voi käyttää yhdessä SAP-järjestelmän kanssa, on rajattu opinnäytetyön ulkopuolelle. Myös teoriaosuuksessa keskitytään niihin malleihin, jotka ovat käytössä SAP-järjestelmässä.

Käytännön osuus koostuu kolmesta osiosta. Ensimmäisessä osassa siirrän tarvittavan testausdatan toimeksiantajan testijärjestelmään ja tarvittaessa muokkaan dataa sellaiseen muotoon, että siitä voidaan tutkia haluttuja asioita. Seuraavaksi tutkin eri mallien toimintaa käytännössä sekä niiden soveltuvuutta ennusteiden arviointiin. Yritän löytää mallin ja prosessin, jonka avulla pystyy arvioimaan ennusteiden onnistumista. Jos löydän sopivan mallin, teen siitä esimerkkitoteutuksen testijärjestelmään. Lopuksi esittelen tutkimukseni tulokset, eli eri mallien toiminnan ja mahdollisen esimerkkitoteutuksen, toimeksiantajalle.

Työn toimeksiantaja on Fujitsu Finland Oy, joka on tietotekniikan palvelu- ja laite-toimittaja, joka toimittaa mm. SAP toiminnanohjausjärjestelmään liittyviä palveluita. Heillä on ollut tarve selvittää, onko tiedonlouhintaa mahdollista hyödyntää tämäntyyppisessä analyysissä. Sain toimeksiannon suoritettua harjoittelujakson siellä.

2 TIEDONLOUHINTA

SAP toimittaa tiedonlouhintametodeja, jotta yritykset voivat tunnistaa mahdollisesti merkittäviä yhteyksiä, trendejä ja malleja, joita ei ole pystytty havaitsemaan perinteisillä metodeilla. Perinteisten analysointimethodien ongelmana on esim. suurten datamäärien käsittelyn vaatima aika ja analysoijien ennakkoodotusten vaikutus tuloksiin. (McDonald ym. 2006, 297 – 298.)

Tiedonlouhinta on joukko metodeita ja tekniikoita, joita käytetään suurten datajoukkojen automaattiseen tai osittain automaattiseen analysointiin. Datasta etsitään tuntemattomia ja piilossa olevia sääntöjä, yhteyksiä ja suuntauksia. Järjestelmät karsivat datamäärää ja antavat tulokseksi vain olennaisimman. (Tufféry 2011, 4.) Tiedonlouhinnan tarkoituksena on löytää kiinnostavaa tietoa suurista datamääristä, jotka on varastoitu tietokantoihin, tietovarastoihin ja muihin tiedon säilytyspaikkoihin. (Han & Kamber 2006, 7.)

Elektronisessa muodossa tallennetun datan määrä on kasvanut valtavasti 1970-luvulta 2000-luvulle. Esimerkiksi vähittäismyynnin, vakuutusalan, terveydenhuollon ja pankkialan dataa kerätään yhä enemmän. Datan varastoinnista tuli helpompaa, kun laskentatehosta tuli halvempaa ja helpommin saavutettavaa. (Prabhu & Venkatesan 2007, 1—2.) Monet tiedonlouhintametodit soveltuvat paremmin suurten kuin pienten datamäärien analysointiin (Berry & Linoff 2011, 3).

Useimmat tiedonlouhintatekniikat ovat olleet olemassa vuosikymmeniä, mutta 1990-luvulta lähtien tiedonlouhinnan käyttö on kasvanut voimakkaasti. Se johtuu muun muassa seuraavista syistä:

- Dataa tuotetaan
- Dataa varastoidaan
- Laskentateho on edullista
- Kaupallisia tiedonlouhintaohjelmistoja on saatavilla. (Berry & Linoff 2011, 6.)

Tiedonlouhinnalla voidaan löytää tuhansia kaavoja tai sääntöjä datajoukosta. Kaikki nämä eivät kuitenkaan ole hyödyllisiä. Tiedonlouhinnan löytämä kaava on hyödyllinen, jos se on helposti ymmärrettävä ihmisen näkökulmasta, toimii testidatalla jollakin tarkkuuden asteella, on mahdollisesti hyödyllinen ja on ennestään tuntematon. Tiedonlouhintaa voidaan käyttää myös jonkin hypoteesin vahvistamiseen. (Han & Kamber 2006, 27—28.)

Kaavojen hyödyllisyyttä voidaan mitata objektiivisesti tai subjektiivisesti. Objektiivisessa tavassa käytetään arviointiin kaavojen rakenteita ja niihin perustuvia tilastoja. Esimerkiksi yhteenkuuluvuusanalyysissä objektiivisia mittareita ovat tuki, luotettavuus ja korotuskerroin, jotka selostetaan luvussa 4.3. Tiedonlouhinnan löytämä kaava saattaa vaikuttaa objektiivisesti kiinnostavalta, mutta osoittautua yleistiedoksi, jolloin siitä ei ole hyötyä. Subjektiivista kiinnostavuutta voidaan mitata käyttäjän uskomusten mukaan. Tällä tavalla arvioituna kaavat ovat kiinnostavia, jos ne ovat odottamattomia tai tarjoavat strategista tietoa, jonka pohjalta käyttäjä voi toimia. Myös se, minkälaista tietoa käyttäjä etsii, vaikuttaa siihen, onko jokin tietty kaava kiinnostava. (Han & Kamber 2006, 28.)

Riippuu tiedonlouhintamallin asetuksista ja käytettävästä metodista, löytääkö tiedonlouhinta kaikki kiinnostavat kaavat ja kuinka paljon turhia kaavoja se havaitsee. On kuitenkin epärealistista odottaa, että tiedonlouhinta löytää kaikki hyödylliset kaavat, eikä lainkaan turhia. (Han & Kamber 2006, 28.)

Tiedonlouhinta ei anna varmaa tietoa datan tulevasta käyttäytymisestä analysoimalla historiallista dataa. Se on ohjaava työkalu, joka tuo näkyville trendejä, jotka voidaan havaita menneisyyttä analysoimalla. (Prabhu & Venkatesan 2007, 14.)

3 TIEDONLOUHINTAPROSESSI

Tiedonlouhintaprosessin vaiheista osa on valinnaisia, mutta yleisesti ottaen prosessin vaiheet ovat:

- tavoitteiden määrittäminen
- olemassa olevan datan luetteloiminen
- datan kerääminen
- datan tutkiminen ja valmisteleva
- populaation jakaminen
- mallien laatiminen ja validoiminen
- mallien ottaminen käyttöön
- käyttäjien kouluttaminen
- mallien toiminnan tarkkaileminen
- mallien parantaminen (Tufféry 2011, 25).

Tavoitteiden määrittäminen

Prosessi aloitetaan valitsemalla aihe, kohderyhmä ja tilastollinen kokonaisuus, joita halutaan tutkia, määrittämällä joitakin välttämättömiä kriteerejä ja ennustettava ilmiö. Prosessin odotetut tulokset on määritettävä ja tätä varten on päätettävä, mihin tuloksia ja louhittua tietoa halutaan käyttää. Tässä vaiheessa päätetään alustavasti, mitä tiedonlouhintatyökaluja käytetään tulosten saavuttamiseksi. Tavoitteiden on oltava tarkat ja niiden pitää johtaa tiettyihin toimiin. Lisäksi niiden on oltava realistiset. (Tufféry 2011, 26.)

Jotta tiedonlouhinta tuottaisi merkityksellistä informaatiota, ennen varsinaista tiedonlouhintaa on datalle tehtävä monia toimenpiteitä, kuten tutkia, valita, siistiä, rikastaa ja muuttaa dataa. Hyvin suunnitellulla tietovarastolla voidaan näistä käsitellä suurin osa. Toisaalta taas tiedonlouhinta parantaa tietovarastointia, kun saadaan selville, mitkä dataelementit ovat merkityksellisimpiä. (Prabhu & Venkatesan 2007, 14.)

Olemassa olevan datan luetteloiminen

Ensimmäinen vaihe datan käsittelyssä on määrittää mitä tutkittavien yksilöiden ominaisuuksiin ja käyttäytymiseen liittyvää dataa on käytettävissä. Datan on oltava hyödyllistä, saavutettavissa olevaa, laillisesti ja teknisesti hyödynnettävissä, luotettavaa ja ajantasaista. Jos tavoitteena on rakentaa ennustava malli, on tarpeen kerätä myös historiallista dataa, jolla malli voidaan opettaa. Mikäli yrityksellä ei ole arkistoituna tarvittavaa dataa, voidaan käyttää näytettä osalta kohderyhmää. Joskus yrityksen data on muodossa, jossa sitä ei voida hyödyntää tiedonlouhinnassa. Silloin tarvittava data on kerättävä ennen tiedon louhimista. (Tufféry 2011, 26–27.)

Datan kerääminen

Seuraavassa vaiheessa luodaan tietokanta, jota käytetään mallien pohjana. Data esitetään tavallisesti taulukossa, jossa yhdellä rivillä on tiettyä yksilöä koskevat tiedot ja sarakkeissa ominaisuudet, jotka kuvaavat yksilöitä. Taulukkoa päivitetään tavallisesti tietyin aikavälein. Joskus dataa on muokattava tässä vaiheessa esim. kokoamalla data tuotetasolta asiakastasolle. Taulukko kootaan tutkimuksen tavoitteiden perusteella, jotta datasta voidaan analysoida haluttuja asioita. Vaihetta varten on tarpeen kerätä dataa useita kuukausia tai jopa vuosia, etenkin kun halutaan käyttää ennustavaa mallia, jotta tulokset ovat tarkkoja. (Tufféry 2011, 27 – 30.)

Datan tutkiminen ja valmisteleminen

Datan valmistelu voidaan jakaa kolmeen eri vaiheeseen, siistimiseen, rikastamiseen ja muuntamiseen. Siistiminen yrittää korjata väärät attribuutit olemassa olevassa datassa. Rikastaminen lisää uusia attribuutteja olemassa olevaan dataan ja muuntaminen muuttaa sen attribuuttien rakennetta niin, että ne soveltuvat tiedonlouhintaan. (Prabhu & Venkatesan 2007, 16.)

Siistimisessä varmistetaan, että datan avaimet ja määritteet ovat yhtenäiset. Vaiheessa tarkastetaan data fyysisten ja loogisten epäjohdonmukaisuuksien varalta kuten tarpeellisten kenttien NULL-arvot ja päättymispäivät, jotka ovat ennen vastaavaa alkamispäivää. Tärkeä osa siistimistä on duplikaattien poistaminen. Jos sama yksilö on esitetty useammin kuin kerran, johtuen esim. kirjoitusvirheestä, on tarpeen poistaa ylimääräiset esiintymät. Toinen tyypillinen virhe on puutteet johdonmukaisuudessa ja täsmällisyydessä. Tämä virhetyyppi on tavallisesti vaikeampi havaita ja se saattaa vaikuttaa paljon tiedonlouhinnan tulokseen. Tällainen virhe on esim. negatiivinen arvo kentässä, jonka arvon tulisi olla aina positiivinen. (Prabhu & Venkatesan 2007, 16—17.)

Jos samassa muuttujassa on virhe useiden yksilöiden kohdalla, muuttuja ei sovellu käytettäväksi. Malli, josta puuttuu yksi muuttuja, on hyödyllisempi, kuin malli, jossa on virheellinen muuttuja. Tämä täytyy ottaa huomioon myös, jos muuttuja ei ole aina saatavilla tai sitä ei välttämättä päivitetä oikein. (Tufféry 2011, 30.)

Datan rikastamisessa lisätään uusia attribuutteja esimerkiksi laskemalla uusia tunnuslukuja tai tuomalla uusia kenttiä ulkopuolisista lähteistä (Prabhu & Venkatesan 2007, 17).

Viimeisessä vaiheessa raakadata käsitellään louhittavaksi sopivaan muotoon. Dataa voidaan käsitellä muotoon, joka sopii paremmin laskemiseen esim. korvaamalla absoluuttiset arvot suhdeluvuilla, muokkaamalla lukuja funktioilla esim. arvojen jakauman tasoittamiseksi, muuttamalla merkkijonot numeroiksi tai yksikkömuunnoksilla. (Tufféry 2011, 30–31.)

Valmistelussa on usein tarpeen yrittää vähentää yksilöitä tutkittavasta populaatiosta, käsiteltäviä muuttujia ja muuttujien kategorioita. Tässä vaiheessa yksilöistä korjataan virheelliset arvot. On myös tavallista käyttää otosta koko populaatiosta, jolloin on varmistettava, että otos vastaa koko populaatiota. Muuttujia voidaan vähentää karsimalla muuttujia, jotka korreloivat jotain toista käytettävää muuttujaa tai eivät ole olennaisia tutkittavan asian kannalta. Lisäksi useampia muuttujia voidaan yhdistää yhteen muuttujaan. Luokittelevaa tai ennustavaa

mallia varten voidaan tutkia 100—200 muuttujaa. Tavallisesti malleissa käytetään jotain seuraavista muuttujamääristä:

- 3—4 yksinkertaiseen malliin
- 5—10 normaaliin malliin
- 11—20 hyvin tarkkaan mallin. (Tufféry 2011, 31—32.)

Tavallisesti mallin yleistettävyys häviää, jos muuttujia on yli 20. Määriä voidaan kuitenkin mukauttaa populaation koon mukaan. Muuttujia, jotka toimivat avainkenttinä, ei voi käyttää ennustamiseen, koska niiden tehtävä on yksilöidä rivit. (Tufféry 2011, 32—33.)

Erillisten tai laadullisten muuttujien kategorioita voidaan vähentää yhdistämällä, jos kategorioita on liikaa, niissä ei ole tarpeeksi yksilöitä tai niillä on sama merkitys tuloksen kannalta. Jatkuvat muuttujat voidaan jakaa kategorioihin. (Tufféry 2011, 32—33.)

Populaation jakaminen

Jos populaation eri jäseniä on tarpeen analysoida eri tavoilla, populaatio voidaan jakaa ryhmiin, joille luodaan omat mallit, ennen kuin ne kootaan yhteen. Jakamista voidaan käyttää vain silloin, kun dataa on niin paljon, että kaikissa ryhmissä on riittävästi yksilöitä. Jakaminen voidaan suorittaa joko asiantuntijan määrittämien sääntöjen avulla tai automaattisesti tilastollisella algoritmilla, kuten päätöspuulla. Populaatio voidaan jakaa esim. sellaisten yleisien ominaisuuksien perusteella, joilla ei ole suoraa yhteyttä ennustettavaan muuttujaan. Ominaisuuksina voidaan käyttää esim. yrityksen kokoa tai toimialaa tai henkilön ikää tai ammattia. Tämä metodi ei aina ole käyttökelpoinen, koska käyttäytyminen ei aina liity yleisiin ominaisuuksiin. Toinen tapa populaation jakamiseen on käyttää toiminnallista dataa, joka on yhteydessä ennustettavaan muuttujaan esim. tuote, johon pisteytys liittyy. Kolmas tapa liittyy saatavilla olevan datan luonteeseen. Vanhoista asiakkaista saattaa olla saatavilla enemmän dataa, kuin mahdollisista uusista asiakkaista, jolloin nämä kaksi on jaettava eri segmentteihin, joille luodaan eri mallit. (Tufféry 2011, 33—34.)

Kaikissa tavoissa jakamisen tulee noudattaa tarkkoja sääntöjä yksilöiden luokiteluun. Lisäksi on varmistettava, että jakamisessa ei käytetä liikaa sääntöjä, segmenttien määrä on rajallinen ja segmentit ovat tasapainoisia, samaa kokoluokkaa ja yhtenäisiä. Aiemmin mainittu yksilöiden, muuttujien ja kategorioiden vähentäminen voidaan suorittaa joko ennen jakamista tai jokaiselle segmentille erikseen. (Tufféry 2011, 34.)

Mallien laatiminen ja validoiminen

Seuraavassa vaiheessa luodaan käytettävä tiedonlouhintamalli. Vaiheessa opetetaan malli tarvittaessa erikseen kaikille edellisen vaiheen segmenteille ja tarkistetaan niiden toiminta datajoukolla, joka on eri kuin opetusvaiheen data. Eri mallien käyttötarkoitukset ja toiminnot kuvataan tarkemmin luvussa 4. Tässä vaiheessa luodaan tavallisesti useampi malli joko samasta tai eri ryhmästä. Tavallisemmin malleja luodaan saman ryhmän metodeilla. Useiden mallien tarkoituksena on optimoida mallin parametrit mukauttamalla valittuja muuttujia, niiden kategorioita ja mallien ominaisuuksia. Voidaan myös luoda eri ryhmien malleja, jotka ajetaan yhdessä parhaan mahdollisen tuloksen saamiseksi. Malleja voidaan tavallisesti vertailla valmiiden tilastollisten indikaattorien perusteella. Matemaattisen tarkkuuden lisäksi mallin valinnassa on otettava huomioon myös esim. mallin lukeminen käyttäjän näkökulmasta tai sen implementointi IT-järjestelmässä. (Tufféry 2011, 25—36.)

Seuraavaksi yhdistetään mallit, jos analysoitava populaatio on jaettu segmentteihin (Tufféry 2011, 36).

Vaiheita datan valmistelusta mallien yhdistämiseen toistetaan tarvittaessa niin kauan, että saadaan tyydyttävä tulos. Dataa muokataan, uusia muuttujia valitaan tai tiedonlouhintamallien parametreja muutetaan halutun tuloksen saavuttamiseksi. Ennen mallin hyväksymistä, se pitäisi tarkastuttaa aiheen asiantuntijoilla, jotka pystyvät arvioimaan tuloksia eri kannalta, kuin tiedonlouhinnan asiantuntijat. He pystyvät havaitsemaan esim. puutteet menneisyystietojen käsitelyssä tai muuttujien valinnassa. (Tufféry 2011, 37.)

Mallien ottaminen käyttöön

Tiedonlouhintamallit implementoidaan käytettävässä järjestelmässä, ennen kuin niiden tuloksia käytetään tai annetaan käyttäjille. Tässä vaiheessa päätetään myös kuinka tarkasti tulokset halutaan esittää, kenellä on pääsy mihinkin tuloksiin ja kuinka usein data on tarpeen päivittää. Tavallinen päivitystiheys on kuukauden välein, mutta tilanteesta riippuen esim. neljännesvuosittainen tai päivittäinen päivitys voi olla parempi. Ennen lopullista käyttöönottoa mallia on hyvä testata useita kuukausia pienellä käyttäjämäärällä. (Tufféry 2011, 37—38.)

Käyttäjien kouluttaminen

Mallin käyttäjille koulutetaan tavoite, työkalujen perusteet, miten ne toimivat, työkalujen rajoitteet, niiden käyttötavat, mitä hyötyä uusista työkaluista on ja miten uudet mallit muuttavat käyttäjien työtapoja (Tufféry 2011, 38).

Mallien toiminnan tarkkaileminen

Jo siinä vaiheessa, kun ongelmaa määritetään, on tarpeen miettiä, miten tuloksia voidaan mitata. Tulosten mittaaminen on tärkeää, jotta voidaan määrittää onko analyysistä ollut toivottua hyötyä ja onko se ensinnäkään onnistunut. Tämä vaihe jää usein vähemmälle huomiolle, koska siitä ei saada suoraa tuottoa. Onnistumisen mittaaminen on tärkeää, koska myös sellaisista tiedonlouhinta-prosesseista, jotka eivät ole onnistuneet halutulla tavalla, voi oppia tulevia prosesseja ajatellen. (Berry & Linoff 2011, 20—23.)

Esimerkiksi markkinointikampanjan kohdentamisen onnistumista voidaan arvioida jakamalla asiakaskunta ryhmiin, joiden tuloksia vertaamalla voidaan selvittää onko kampanja kohdistettu oikein ja onko kampanjasta ollut hyötyä. Vertailtavat ryhmät ovat:

- Kohderyhmä, joka on mallin valitsema, ja jolle kampanja kohdistetaan

- Vertailuryhmä, jota malli ei ole valinnut, mutta jolle kuitenkin kohdistetaan kampanja
 - Ryhmä, jota malli ei ole valinnut ja jolle ei kohdisteta kampanjaa
 - Ryhmä, jonka malli on valinnut, mutta jolle ei kohdisteta kampanjaa.
- (Berry & Linoff 2011, 21.)

Toinen tapa tarkkailla mallia on jatkuva. Sitä käytetään malleihin, jotka ovat käytössä jatkuvasti. Tuloksia kerätään esim. taulukkoon, minkä perusteella arvioidaan mallin toiminnan oikeellisuutta ja tarkkuutta. (Tufféry 2011, 38—40.)

Mallien parantaminen

Prosessin viimeisessä vaiheessa käytössä olevaa mallia parannetaan. Uuden parannetun mallin luominen voidaan aloittaa mistä tahansa vaiheesta tavoitteiden määrittämisen jälkeen. Prosessi on hyvin iteratiivinen ja samoja vaiheita toistetaan koko mallin elinkaaren ajan. (Tufféry 2011, 40—41.)

4 TIEDONLOUHINTAMETODIT

Tiedonlouhintaan on tarjolla useita eri metodeita. Eri tekniikat tyypillisesti vaativat tietynlaisen datarakenteen ja niiden taustalla on erityyppiset algoritmit. (Prabhu & Venkatesan 2007, 23.) Tiedonlouhintametodit voidaan jakaa kuvaaviin ja ennustaviin metodeihin. Kuvaavissa malleissa datan vähentämiseen, tiivistämiseen ja ryhmittelyyn ei käytetä tiettyä muuttujaa. Ennustavissa metodeissa, joiden tarkoitus on selittää dataa, on muuttujia, jota yritetään selittää. Kuvaavia metodeja ovat esimerkiksi faktorianalyysi, klusterianalyysit ja yhteyksien etsiminen. Ennustavia metodeja ovat esimerkiksi päätöspuut ja neuroverkot. (Tufféry 2011, 167—169.)

Käsittelen tässä opinnäytetyössä ainoastaan ne tiedonlouhintamallit, jotka ovat käytössä SAP-järjestelmässä.

SAP:n tarjoamat tiedonlouhintametodit jaetaan kolmeen kategoriaan, luokitteluun, ryvästyksen ja yhteenkuuluvuusanalyysiin. Malleista päätöspuut ja pisteytysmallit kuuluvat luokitteluun. (McDonald ym. 2006, 301.)

Toiminnallisuuden mukaan metodit voi jakaa analyttisiin ja ennustaviin metodeihin. Analyttisiä metodeja, jotka laskevat tulokset ilman historiallista dataa ovat ABC-luokitus, yhteenkuuluvuusanalyysi ja painotetut pisteytystaulukot. Ennustavia metodeja ovat päätöspuut ja regressioanalyysi. Ryvästystä voidaan käyttää sekä analyttisenä että ennustavana metodina. (SAP Help Portal 2013a, 2013d, 2013e, 2013j.)

Mallin opetuksessa, opetetun mallin arvioinnissa ja lopullisen ennusteen luomisessa käytetään eri dataa. Opetuksessa käytetään sellaista datajoukkoa, joka edustaa mahdollisimman hyvin dataa, jonka analysointiin mallia aiotaan käyttää. Sekä arvioinnissa että ennustamisessa käytetään historiallista dataa, mutta eri vaiheissa pitäisi käyttää eri datajoukkoa. Arvioinnin tarkoituksena on tarkastaa opetettu malli, ennen kuin sitä käytetään ennusteen luomiseen. (McDonald ym. 2006, 301 - 302.)

Opetukseen käytettävän datamäärään tulisi olla riittävän suuri, jotta malli on tarkka, muttei niin suuri, että se vaikuttaa suorituskyykyyn. Opetettu malli arvioidaan, jotta voidaan määrittää opetuksen onnistuminen. Kun malli on opetettu ja arvioitu, sitä voidaan käyttää ennustamiseen. Ennusteen tulos riippuu valitusta metodista. (McDonald ym. 2006, 302.)

SAP-järjestelmässä mallit luodaan transaktiolla RSDMWB ja niitä käytetään luomalla prosessi analyysiprosessisuunnittelijassa, jonka transaktiokoodi on RSANWB (McDonald ym. 2006, 301).

4.1 ABC-luokitus

ABC-luokitus (ABC Classification) on analyttinen malli, jota käytetään luokittelemaan objekteja kuten asiakkaat, tuotteet tai työntekijät perustuen tiettyyn tunnuslukuun esim. tuotot. Yritys voi esimerkiksi luokitella asiakkaansa luokkiin A, B ja C sen mukaan, kuinka paljon myyntituloa kukin asiakas tuo yritykseen. Luokittelu voidaan suorittaa joko luokitteluperusteen, eli luokitteluun käytettävän tunnusluvun, tai luokiteltavien objektien lukumäärän mukaan. Molemmilla tavoilla voidaan käyttää joko absoluuttisia arvoja tai kumulatiivisia prosenttiosuuksia. Luokkien raja-arvot siis asetetaan luokitteluperusteen absoluuttisina arvoina, luokitteluperusteen arvojen kumulatiivisen kertymän mukaan, luokiteltavien objektien lukumäärän mukaan tai lukumäärän kumulatiivisen kertymän mukaan. (SAP Help Portal 2013a, 2013c.)

Luokittelua käytetään jakamaan data ennalta määriteltyihin luokkiin. Tiedonlouhinnassa luokittelua on käytetty esimerkiksi luokittelemaan trendejä rahoitusmarkkinoilla tai tunnistamaan automaattisesti kiinnostavat objektit suurista tietokannoista. (Prabhu & Venkatesan 2007, 23.)

4.2 Pisteytys

Pisteytystä (Scoring) käytetään ennustamaan numeerisia arvoja historiallisen datan perusteella. Pisteytystä voidaan käyttää esimerkiksi uuden tuotteen tuot-

tojen ennustamiseen, jos niiden voidaan olettaa korreloivan samalla tavalla, kuin jonkin vanhan tuotteen tuotot, esimerkiksi kohderyhmän vuositulojen ja talouden koon kanssa. (SAP Help Portal 2013i.) Pisteytys on alun perin kehitetty asiakkuudenhallintajärjestelmää varten, mutta niille on kehitetty myös muita sovelluksia. Pisteytyksessä on valittavana kaksi eri regressiotyyppiä:

- Lineaarinen regressio tilanteisiin, joissa määritteillä on lineaarinen riippuvuus ennustetunnusluvusta
- Epälineaarinen regressio tilanteisiin, joissa suhde ei ole lineaarinen (McDonald ym. 2006, 305.)

Lineaarisessa regressiossa pisteytys opetetaan suorittamalla historialliseen dataan lineaarisen regression algoritmeja. Järjestelmä luo erilliset lineaariset funktiot jokaiselle yhdistelmälle eri arvoista opetuslähteessä. Jos kahdella erillisellä mallikentällä on kummallakin viisi ominaisuutta, suoritetaan 25 lineaarista regressioanalyysia. Tämä täytyy ottaa huomioon suorituskyvyn kannalta, koska pisteytysmalli luo karteesisen tulon. (McDonald ym. 2006, 305.)

Epälineaarinen regressio käyttää pisteytykseen samantapaista menetelmää, kuin lineaarinen regressio, mutta siinä ei oleteta olevan lineaarista yhteyttä kenttien arvojen välillä. (McDonald ym. 2006, 305.)

4.3 Painotettu pisteytystaulukko

Painotetuilla pisteytystaulukoilla (Weighted score tables) ei pisteytykseen käytetä lainkaan historiallista dataa. Käyttäjä antaa manuaalisesti pistearvot ja painotuskertoimet ominaisuuksien arvoille. Tätä metodia voidaan käyttää esimerkiksi määrittämään, kuinka todennäköisesti asiakas ostaa uuden tuotteen. Esimerkiksi asiakkaan iälle ja asuinympäristölle annetaan painotukset sen mukaan, miten arvojen oletetaan vaikuttavan tuotteen ostoon. Jos painotusten määrittäminen on onnistunut, saatu pistemäärä kertoo, onko todennäköistä, että asiakas ostaa uuden tuotteen. (McDonald ym. 2006, 305.)

Painotetun pisteytystaulukon mallissa käyttäjä valitsee jokaiselle kentälle erikseen painotuksen. Sen lisäksi käyttäjä antaa kenttien arvoille osapainotukset. Nämä kaksi määrittävät yhdessä sen, kuinka suuri lopullinen painotus on. Se lasketaan seuraavalla tavalla: $\text{Pisteet}(\text{kenttä1}, \text{kenttä2}, \dots) = \text{painotus1} * \text{osapainotus1}(\text{kenttä1}) + \text{painotus2} * \text{osapainotus2}(\text{kenttä2}) + \dots$ (SAP Help Portal 2013k.)

Mallin kentät voivat olla tyypiltään joko jatkuvia tai epäjatkuvia. Jatkuvilla arvoilla osapainotukset annetaan raja-arvoille ja arvot niiden välillä voidaan painottaa joko raja-arvon osapainotuksen mukaan tai käyttäen lineaarista interpolointia. Epäjatkuvilla kentillä osapainotus annetaan erikseen kaikille arvoille. (SAP Help Portal 2013k.)

4.4 Yhteenkuuluvuusanalyysi

Yhteenkuuluvuusanalyysi (Association analysis) määrittää asioiden riippuvuutta toisistaan. Sitä käytetään paljon vähittäiskaupassa määrittämässä mitä tuotteita myydään yhdessä, mutta myös muille toimialoille on käytännöllisiä sovelluksia. (McDonald ym. 2006, 310.)

Analyysin tuloksia voidaan käyttää esimerkiksi yhdessä myytävien tuotteiden sijoittamiseen kuvastossa, myymälässä tai nettikaupassa tai kohdistamaan markkinointikampanjoita asiakkaille, jotka tuotteen A ostettuaan, todennäköisesti ostavat myös tuotteen C. (SAP Help Portal 2013b.)

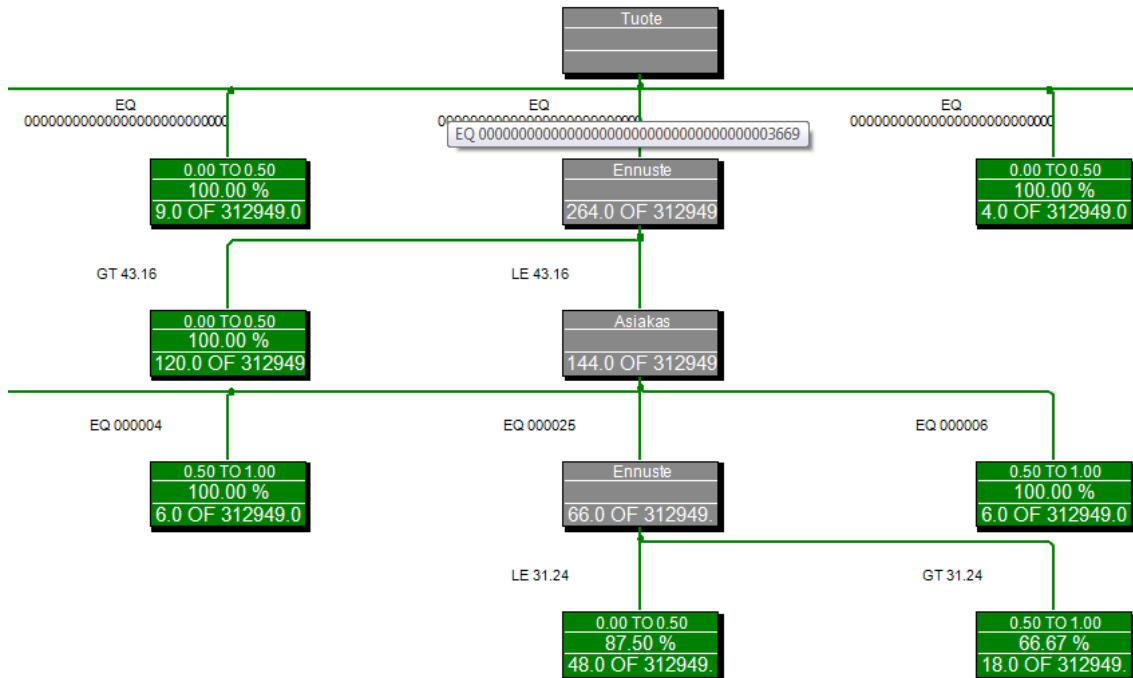
Yhteenkuuluvuusanalyysin konfiguroinnissa ja tulosten lukemisessa on kolme tärkeää termiä tuki (support), luotettavuus (confidence) ja korotuskerroin (lift). Tuki kertoo prosenteissa kuinka usein tietty yhdistelmä esiintyy. Luotettavuus on prosentuaalinen todennäköisyys sille, että sidonnainen rivi esiintyy datajoukossa, jossa pääryvi on esiintynyt. Korotuskerroin mittaa yhteenkuuluvuusanalyysin onnistuneisuutta laskemalla tulosten, joilla on yhteenkuuluvuussääntö ja joilla ei ole, suhde. Sitä käytetään poistamaan tuloksista ne yhdistelmät, jotka eivät ilmennä todellista yhteenkuuluvuutta, mutta esiintyvät silti usein yhdessä. (McDonald ym. 2006, 310.)

Mallin konfiguroinnissa voidaan valita kullekin tunnusluvulle raja-arvot, joiden avulla tuloksista karsitaan epäolennaisuudet. Tulosten karsiminen parantaa suorituskkyä ja tulosten luettavuutta. (McDonald ym. 2006, 310.)

4.5 Päättöspuu

Päätöspuita (Decision tree) käytetään niin, että ne opetetaan historiallisella datalla, jolla luodaan säännöt. Näitä sääntöjä käytetään ennustamaan tulevaisuutta soveltamalla niitä uuteen dataan. (SAP Help Portal 2013f.)

Päätöspuut luokittelevat datan todennäköisyshierarkioihin. Niitä käytetään esim. asiakasuskollisuuden mittaamiseen. Eri hierarkiatasot edustavat ominaisuuden määrittä (kuten asiakkaan ammatti, ikä tai tulot) ja siihen liittyvää todennäköisyyttä päätyä tiettyyn tulokseen. Menneisyystietoihin perustuvista laskelmista voidaan tehdä yleistyksiä koskemaan esim. uutta asiakasta tai eri asiakasryhmää. (McDonald ym. 2006, 303.)



Kuva 1. Esimerkki päätöspuusta.

Kuvassa 1 on esimerkkinä osa SAP-järjestelmän tuottamasta päätöspuusta. Vastaavan kuvan saa tulokseksi opetusvaiheen jälkeen. Puussa on pyritty tutkimaan todennäköisintä eroa toteuman ja ennusteen välillä, joka lukuarvona on jaettu neljään eri väliin, johon tulos voi sijoittua, esim. 0,50 – 1,00. Ylimmäksi hierarkiatasoksi on arvioitu tuote, ja ylimmän rivin reunimmaisat solmut kuvaavat tuotteita, jotka on voitu sijoittaa tiettyyn jaksoon 100 %:lla tarkkuudella pelkän tuotteen perusteella. Tuote 3669 tarvitsee sopivimman jakson valintaan lisäksi ennusteen koon, asiakkaan ja osalla asiakkaista vielä uudelleen jaottelun ennusteen koon mukaan. Tällä tavalla luotujen sääntöjen mukaan, voidaan ennustaa tuote-asiakas-ennuste—yhdistelmien käyttäytymistä tulevaisuudessa.

4.6 Ryvästys

Ryvästys (Clustering) kokoaa datan segmentteihin eri ominaisuuksien yhteyksiin mukaan. Ryvästyksen löydöksiä voidaan käyttää esimerkiksi asiakkaiden tai markkinoiden segmentointiin ja näin ollen markkinoinnin kohdistamiseen sopivimmille asiakkaille. (McDonald ym. 2006, 309.)

Ryvästys jakaa datan niin, että rivit, joilla on samankaltainen sisältö kuuluvat samaan klusteriin ja rivit, joiden sisältö on erilainen, kuuluvat eri klustereihin. (McDonald ym. 2006, 309.) Ryvästysmalli opetetaan historiallisella datalla. Jos mallia halutaan käyttää ennustamiseen, tulokset lasketaan käyttäen uutta datajoukkoa, joka haetaan jakaa segmentteihin samoin, kuin alkuperäinen data on jakautunut. Jos mallia halutaan käyttää analysointitarkoitukseen, opetettua mallia käytetään samaan datajoukkoon, jolla se on opetettu. Tällöin tulokseksi saadaan kyseisen datajoukon yksilöille sopivimmat klusterit. (SAP Help Portal 2013d.)

5 MYNNINSUUNNITTELU SAP APO DP:SSÄ

SAP APO, eli Advanced Planner and Optimizer, on toimitusketjun hallintaan tarkoitettu järjestelmä. SAP APO oli alun perin tarkoitettu suunnitteluun, joka keskittyy tulevaisuuteen. Sen tarkoitus oli tuottaa suosituksia, mutta erona toimitusketjun suunnitteluun, se ei voinut toteuttaa niitä. Myöhemmin SAP APO on laajentunut sisältämään myös muita toimitusketjun toiminnallisuuksia, jotka puuttuvat SAP ERP:stä. (Snapp 2010, 32—33.)

SAP APO sisältää moduulit:

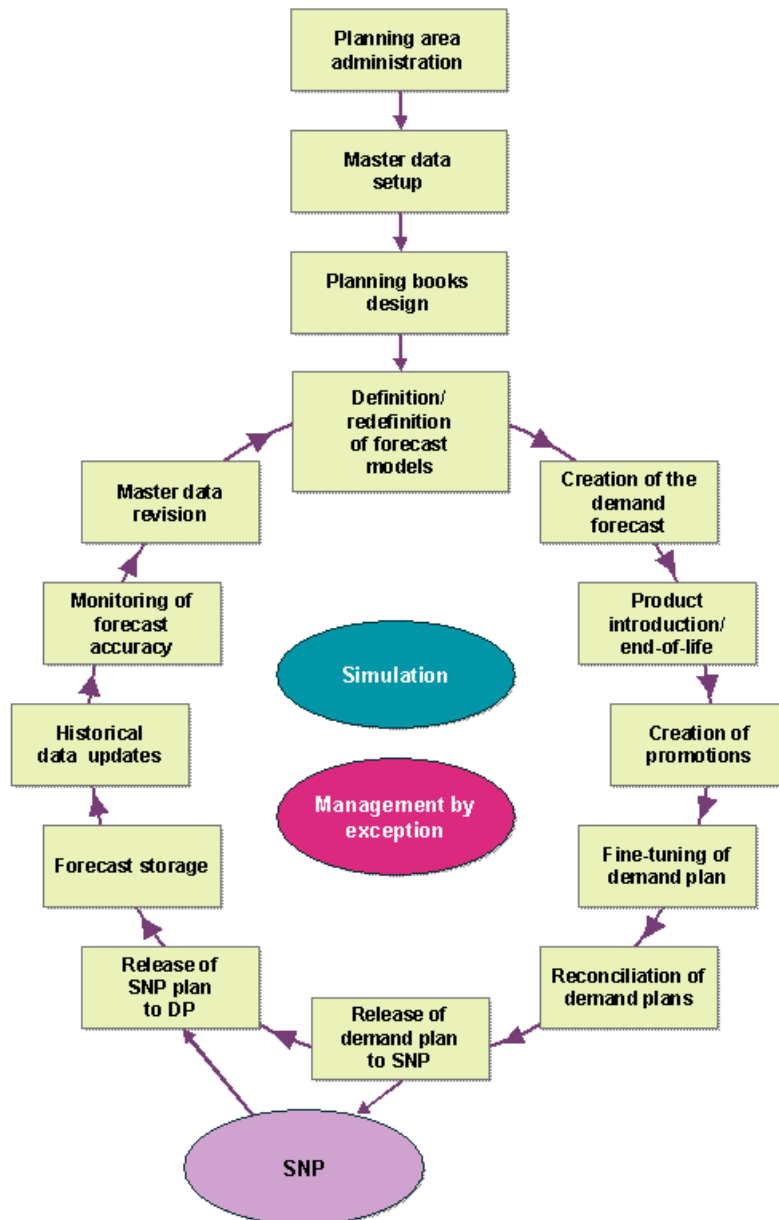
- Demand Planning (DP),
- Supply Network Planning (SNP),
- Production Planning & Detailed Scheduling (PP/DS),
- Available-to-Promise (ATP),
- Transportation Planning & Vehicle Scheduling (TP/VS) (Dickersbach 2009, 11).

Moduuleista DP on tarkoitettu myynninsuunnitteluun. Komponentilla pystyy ottamaan huomioon useita eri tekijöitä, jotka vaikuttavat kysyntään. DP on tehokas ja joustava työkalu, joka tukee myynninsuunnitteluprosessia. Prosessin tuloksena on myynninsuunnitelma. (SAP Help Portal 2013g.)

SAP DP koostuu kahdesta tärkeästä komponentista, jotka ovat Data Warehouse Workbench ja itse suunnitteluosa. Tietovaraston tehtävä on määrittää tarvittavat datarakenteet ja siirtosäännöt datan tuomiseksi malliin. Myynninsuunnittelu käyttää suuria määriä historiallista dataa, joten sillä on enemmän dataa kuin millään muulla SCM sovelluksella. (Snapp 2010, 61.)

Asiakkaiden kasvavat odotukset, tiukemmat säädökset, muuttuvat markkinat ja internetin vaikutukset, ovat lähes kaikilla aloilla johtaneet siihen, että yritysten on pitänyt uudistaa kysynnän ennustaminen ja hallinta. Kysynnän suunnittelu on toimitusketjun hallinnan ehkä kriittisin vaihe. (Snapp 2010, 59.)

Myynnin suunnitelma tuotetaan yleensä 1—5 vuotta tulevaisuuteen ja tavallisesti suunnitelma luodaan joko viikko- tai kuukausitasolla. (Dickersbach 2009, 10—12.)



Kuvio 1. Myynninsuunnitteluprosessi (SAP Help Portal 2013h).

Kuviossa 1 on kuvattu myynninsuunnittelun koko prosessi. Prosessi yleensä suoritetaan kuviossa esitetyssä järjestyksessä, mutta koska prosessi on esitetty kehänä eikä lineaarisesti, voidaan järjestystä joiltain osin tarvittaessa muuttaa. Kolme ylintä vaihetta valmistelevat myynninsuunnittelua. Suunnittelualue (Plan-

ning area) on myynninsuunnittelun kaikkien toimintojen pohja APO-järjestelmässä. Siinä määritetään parametrit kaikille myynninsuunnittelun toiminnoille. Kun suunnittelualue on valmis, perustietoasetuksissa määritetään suunniteltavien objektien perustiedot. Yhdelle suunnittelualueelle voidaan luoda useita suunnittelukansioita (Planning book) esimerkiksi eri käyttäjille, eri hierarkiatasojille ja erimittaisille suunnittelujaksoille. (SAP Help Portal 2013h.)

Seuraavassa vaiheessa etsitään sopiva malli, jonka avulla objektin myyntiä voidaan ennustaa. Kun on valittu paras tapa ennustaa myyntiä, voidaan ennuste luoda ja tehdä sille joitakin tarkennuksia, jotka kuvataan seuraavassa neljässä vaiheessa. Ennen SNP:hen vapauttamista, eri objektien ennusteet on yhdistettävä. Kuviossa alimpana oleva SNP kuvaa DP:n yhteyttä SNP:hen. (SAP Help Portal 2013h.)

Eri aikoina luodut ennusteet säilytetään. Seuraavaksi päivitetään historiallinen data. Ennusteiden onnistumista voidaan arvioida esim. vertaamalla niitä toteutuneeseen myyntiin. Perustietoja päivitetään tarpeen mukaan ja prosessi kierroksen ulkopuolella olevia vaiheita lukuun ottamatta toistetaan uusien ennustettavien objektien kohdalla. (SAP Help Portal 2013h.) Tässä opinnäytetyössä keskitytään vaiheeseen ”Monitoring of forecast accuracy”, eli ennusteiden tarkkuuden valvontaan.

Myynnin suunnittelu voidaan ajaa taustalla tai käyttää vuorovaikutteista myynninsuunnittelua. Tausta-ajo on tehokkain tapa prosessoinnin kannalta. (Snapp 2010, 63.) Suurin osa suunnittelutoiminnoista suoritetaan taustalla, joko siksi, että ne suoritetaan kausittain tai niitä käytetään suurten datamäärien prosessointiin (Dickersbach 2009, 77).

Vuorovaikutteinen suunnittelu on kuitenkin hyödyllinen esimerkiksi silloin, kun suunnittelija haluaa nähdä muutoksien vaikutuksen ennusteeseen tai yrittää valmistella uuden mallin käyttöönottoa, koska se näyttää muutosten vaikutukset nopeasti. Vuorovaikutteinen suunnittelu sopii parhaiten tulosten testaamiseen rajatulla datajoukolla niin, että suunnittelija tutkii muutosten vaikutusta ennusteeseen. (Snapp 2010, 63—64.)

6 ENNUSTEVERSIOIDEN ANALYSOINTI

6.1 Tehtävän määrittely

Tehtävänäni on tutkia, onko mahdollista käyttää SAP-järjestelmän tarjoamia tiedonloughintatyökaluja ennusteverversioiden analysoinnissa. Ennusteet on luotu APO:n myynninsuunnittelussa ja eri versiot ovat eri viikoilla tallennettuja ennusteita. Niitä toteumaan vertaamalla voidaan arvioida ennusteiden onnistumista. Minun tehtäväni on selvittää, onko tämä arviointi mahdollista tehdä tiedonloughintamalleja hyödyntäen.

Toimeksiantaja on kiinnostunut ratkaisusta ennusteiden onnistumisen arviointiin tuote- ja asiakaskohtaisesti. Vaikuttaisi siltä, että nämä kaksi objektia ovat sen verran samankaltaisia, että niiden analyysit voidaan toteuttaa samanlaisilla prosesseilla, muuttamalla ainoastaan prosessiin tuotavaa dataa ja joitakin parametreja. Parhaassa tapauksessa dataa tarvitsisi käsitellä mahdollisimman vähän ennen tiedonloughintaa.

Tavoitteena on saada tulos, joka esittää, miten hyvin arvioidun objektin ennustaminen on onnistunut. Mallia halutaan käyttää yksinomaan muualla laadittujen ennusteiden arviointiin. Tavoitteena ei ole ennustaa mitään tai etsiä syitä ennusteiden onnistumiselle tai epäonnistumiselle, ainoastaan määrittää ne tuotteet tai asiakkaat, joiden kohdalla ennuste on onnistunut hyvin tai huonosti.

Tätä varten tarvitsee luoda analyysiprosessi, joka laskee tunnusluvun tai tunnuslukuja, joiden kautta voidaan arvioida ennusteen onnistumista yhden tai useamman tiedonloughintamallin avulla.

6.2 Valmistelut

Sain tutkimusta varten testidataa CSV-tiedostoissa, mistä data piti ladata järjestelmään. Testidata sisälsi viikoilla 50.2012, 52.2012, 02.2013 ja 04.2013 tallennetut ennusteet kyseisestä viikosta yksi viikko taaksepäin ja 54 viikkoa eteen-

päin, toteuman viikoille 50.2012 - 11.2013, sekä tuotekohtaisen suunnittelijatie-
don.

Ennusteveriot sisälsivät seuraavat ominaisuudet:

- tuote,
- versiotunniste, joka kertoo millä viikolla ennuste on luotu,
- asiakasryhmät hierarkiatasoilla 1 - 3, joista taso kolme edustaa yksittäis-
tä asiakasta ja tasot 1 ja 2 asiakasryhmiä,
- maa, jossa asiakas sijaitsee,
- myyntiorganisaatio, johon asiakas kuuluu,
- viikko, jota kukin tunnusluku koskee,
- toimipiste, josta tuotetta on tilattu,
- kampanja, johon on tallennettu kyseiselle viikolle suunnitellusta kampan-
jasta, ja sen yksikkö,
- ennuste, joka on luotu myynninsuunnittelussa, ja sen yksikkö.

Toteumadata sisälsi seuraavat ominaisuudet:

- maa,
- toimipiste,
- asiakasryhmät hierarkiatasoilla 1 – 3,
- tuote,
- myyntiorganisaatio,
- viikko, jota kyseinen toteuma koskee,
- toteuma, joka kuvaa kunkin ominaisuusyhdistelmän myyntiä, ja sen yk-
sikkö.

Tuotekohtainen suunnittelija sisälsi listan tuotteista ja niiden suunnittelijoista.

Käytin tietolataukseen osittain valmiita ja osittain itse tekemiäni info-objekteja, jotka ovat SAP tietovaraston pienimpiä tallennusyksiköitä. Itse tekemiäni info-objekteja voin tarvittaessa muokata, sen häiritsemättä muita testijärjestelmän käyttäjiä. Itse tekemäni objektit olivat asiakasryhmiä lukuun ottamatta kopioita

sä on opetusvaihe tai yhteenkuuluvuusanalyysi. Tietolähteitä ovat mm. perustietomääritteet, infoProviderit, kyselyt, tietokantataulut ja tiedostot. Muunnokset ovat erilaisia työkaluja datan käsittelyyn. Niiden avulla voidaan esim. yhdistää eri lähteistä peräisin oleva data, järjestää tai suodattaa dataa, suorittaa erilaisia laskutoimituksia tai koota dataa. Erilaisia tietokohteita ovat esim. perustietomääritteet, DataStore-objektit, CSV-tiedosto tai CRM-määritteet.

Ensimmäinen vaihe on tietolähteiden valinta. Prosessit useimmiten tarvitsevat sekä toteumaa että ennusteversioita, joten prosesseissa on yleensä kaksi tietolähdettä. Tuotteen navigointimääritteenä suunnittelijatietoon pääsee tarvitessa käsiksi kummankin kuution kautta. Samalla valitaan kentät, jotka prosessiin tuodaan. Ne riippuvat siitä, mitä prosessissa analysoidaan. Kyseessä olevan ongelman ratkaisua varten prosessiin on tarpeen tuoda analysoitavat objektit eli tuote tai asiakas, sekä toteuma, ennuste ja niiden yksiköt, sekä versio ja viikko, jota kyseinen tunnusluku koskee.

Tietolähteen lisäksi prosessissa tarvitaan joitakin muuntoja, joiden avulla data käsitellään. Kahdesta eri tietolähteestä peräisin oleva data yhdistetään niin, että kukin ennustettu viikko yhdistetään sitä vastaavaan toteumaan, jotta niiden ero voidaan laskea. Mukaan otetaan ainoastaan ne rivit, jotka löytyvät kummastakin tietolähteestä. Tämän ansiosta mukaan otettavien ennusteiden ja toteumien pitäisi olla toisiaan vastaavat. Samalla karsitaan esim. niiden viikkojen ennusteet, jotka ovat edelleen tulevaisuudessa, eikä niille ole vielä toteumaa.

Kun toteuma ja ennuste on yhdistetty, voidaan laskea niiden välinen ero hyödyntäen analyysiprosessisuunnittelijan kaava-muuntoa. Ero voidaan laskea esim. toteuman ja ennusteen välisenä absoluuttisena arvona, suhteuttamalla toteuman ja ennusteen välinen ero toteumalla tai laskemalla toteuman mukaan painotettu keskiarvo eroista. Eri versioille ja viikoille laskettu ero on tarpeen koota niin, että arvioitavalle objektille saadaan yksi arvo, jonka mukaan sitä arvioidaan. Kokoaminen voidaan tehdä esimerkiksi laskemalla eri arvojen keskiarvo tai laskemalla arvot yhteen.

6.4 Mallin valinta

Vaihtoehtoisia malleja ratkaisua varten on siis ABC-luokitus, yhteenkuuluvuus-analyysi, pisteytys, painotetut pisteytystaulukot, päätöspuu, ryvästys tai jonkinlainen yhdistelmä näistä malleista.

Yhteenkuuluvuusanalyysin sopimattomuus oli pääteltävissä sen toiminnallisuuden perusteella. Metodilla on tarkoitus etsiä objekteja, jotka usein esiintyvät yhdessä. Sen vuoksi se ei sovellu ennusteiden onnistumisen arviointiin. Tulokset esitetään muodossa, joka listaa päärivit ja kullekin sidonnaiset rivit ja niiden tuen, luotettavuuden ja korotuskertoimen. Yhteenkuuluvuusanalyysin tulosten hyödyntämiseen ja esittämiseen muualla kuin CRM- eli asiakkuudenhallintamoduulissa ei myöskään ole tarjolla käytännöllisiä toimintoja.

Ryvästyksessä on tarkoitus yhdistellä valittujen ominaisuuksien mukaan samantyyppisiä objekteja samoihin ryhmiin. Jos mallista haluaa tulokseksi ainoastaan tietyn datajoukon ryhmittelyn, voidaan malli opettaa, ja laskea tulokset samalla datajoukolla, jolloin tulokseksi saadaan jokaiselle objektille sopivimmat klusterit. Kokeilin ratkaisua, jossa tuotteet jaettiin ryhmään niiden ennusteen, toteuman ja toteumalla suhteutetun keskipoikkeaman mukaan. Tulokset eivät kuitenkaan kertoneet varsinaisesti mitään ennusteen onnistumisesta. Tällä ratkaisulla voidaan löytää esimerkiksi yllättäviä yhtenevyyksiä eri tuotteiden käyttäytymisestä.

Päätöspuu ja pisteytys ovat ensisijaisesti tarkoitettu ennustamaan tulevaa menneisyysdatan perusteella. Opetusvaiheen tulokset ovat sellaisessa muodossa, että ne eivät kummassakaan ole sellaisenaan hyödynnettävissä. Pisteytyksen tulokset esitetään kaavioina ja indikaattoreina, jotka eivät kerro yksittäisistä objekteista mitään. Päätöspuun tulokset annetaan sääntöinä, jotka listaavat tulokseen vaikuttavat tasot sekä tuloksen, johon yhdistelmä johtaa ja sen todennäköisyyden.

Päätöspuun ja pisteytyksen tulokset lasketaan niin, että opetettua mallia käytetään ennustamaan valittua objektia toisten objektien perusteella. Jos siis tulokset lasketaan samalla datalla, kuin opetukseen on käytetty, malli yrittää ennus-

taa menneisyyttä. Sillä ei siis saa tulosta, joka kertoisi ennusteen onnistumisesta.

ABC-luokitus ja painotetut pisteytystaulukot ovat kumpikin analyyttisiä malleja, joten ne eivät yritä ennustaa mitään. ABC-luokitus on asetuksiltaan ja toiminnaltaan malleista yksinkertaisin. Se antaa tulokset helposti ymmärrettävässä muodossa ja ne antavat selkeän vastauksen onko ennuste onnistunut hyvin vai huonosti. Jos prosessista saadaan yksi tunnusluku, jonka perusteella ennusteen onnistumista tietyn objektin tai objektiyhdistelmän kohdalla voidaan arvioida, ABC-luokituksella saadaan selkeät tulokset. Esimerkiksi arvioitaessa ennusteen onnistumista tietyn tuotteen tai asiakkaan kohdalla, voidaan laskea tuote- tai asiakaskohtainen keskimääräinen ero ennusteen ja toteuman välillä ja luokitella tuotteet tai asiakkaat sen perusteella.

Jos arvioinnissa halutaan ottaa huomioon ominaisuus tai enemmän kuin yksi tunnusluku, ABC-luokitus yksinään ei käy. Monimutkaisemman analyysin suorittamisessa voi hyödyntää esim. painotettuja pisteytystaulukoita. Siinä annetaan manuaalisesti pistemäärät valittujen kenttien eri arvoille. Jos siis analyysi halutaan tehdä muillakin perusteilla, kuin keskimääräinen ero ennusteen ja toteuman välillä, voidaan määrittää pisteet näille eri perusteille.

Näissä esimerkeissä kuitenkin riittää, että objekteja arvioidaan yhden tunnusluvun perusteella, joten valitsin lähempään tarkasteluun ABC-luokituksen.

6.5 ABC-luokituksen tutkiminen

Tutkimusvaiheessa, kun testasin useita eri prosessivaihtoehtoja ja laskutapoja, käytin tietokohteina CSV-tiedostoja. Se oli paras vaihtoehto sen vuoksi, että tiedoston rakenne määräytyy aina sinne syötettävän datan mukaan, kun taas muissa datan tallennusmuodoissa tulokset tallennetaan info-objekteihin. Lisäksi CSV-tiedostoista pystyy tallentamaan useita eri versioita, joita oli helppo vertailla keskenään.

Päätin, että etsisin ensin sopivimman analyysin tuotteelle, minkä jälkeen voisin testata samanlaisen analyysin toimivuutta asiakkaalle ja lähteä etsimään sille sopivia parametreja.

Arviointiin käytettävän tunnusluvun laskemisessa oli useita eri vaihtoehtoja, joista piti valita parhaiten ennusteen onnistumista ilmentävä.

Aloitin sopivan prosessin etsimisen testaamalla prosessia, joka käytti analyysiin toteuman ja ennusteen välistä eroa. Prosessi toi dataa toteuma- ja ennustekuutiosta. Molemmista kuutioista tuotiin tuotenumero ja viikko, jota tunnusluku koskee, ja niitä käytettiin avaimina yhdistämisessä. Niiden lisäksi tuotiin toteuma sekä ennuste ja sen versio. Yhdistämisen jälkeen laskettiin kullakin rivillä itseisarvo toteuman ja ennusteen erotuksesta. Sen jälkeen analyysiprosessi laski koontityökalua hyödyntäen ensin viikkokohtaisen keskiarvon eri versioiden eroista ja sitten näistä keskiarvon kullekin tuotteelle.

Jo tässä vaiheessa näki, että erot noudattivat hyvin tarkasti samaa suuruusjärjestystä, kuin toteumat, eli niillä tuotteilla, joilla oli suurin toteuma, oli myös suurin ero toteuman ja ennusteen välillä. Tämän vuoksi absoluuttinen ero toteuman ja ennusteen välillä ei ole vertailukelpoinen, eri tuotteiden välillä, jos niiden myynti poikkeaa toisistaan paljon. Tätä tunnuslukua voidaan hyödyntää esimerkiksi tuotteiden luokitteluun sen perusteella, mitkä niistä tulevat todennäköisimmin aiheuttamaan suurimman hävikin.

Toinen tapa määrittää ennusteen onnistumista on suhteuttaa edeltävässä prosessissa laskettu ero tuotteen toteumalla, josta saadaan suhdeluku. Tämä suhdeluku on vertailukelpoinen eri tuotteiden välillä, koska niillä tuotteilla, joilla on suurin ero suuren volyymin vuoksi, on myös suuri jakaja. Tässä laskutavassa oli kuitenkin erikseen otettava huomioon tuotteet, joiden toteuma arvioitavalta ajalta oli nolla. Käytin näissä tapauksissa jakajana yleensä arvoa sata, koska muuten arvot jäivät liian suuriksi ja vääristivät tuloksia. Näitä tuotteita ei ollut testitavassa kuin pari, joten ongelma oli pieni.

Näillä arvoilla aloin etsiä sopivia asetuksia ABC-luokitukselle. Mallille ei tarvitse määrittää muita kenttiä kuin tuote, joka siis on luokiteltava objekti, ja ero, joka

on luokitteluperuste. Mallin voi asettaa luokittelemaan objektit joko luokitteluperusteen tai luokiteltavien objektien lukumäärän mukaan ja molemmilla tavoilla voi valita joko kumulatiiviset tai absoluuttiset arvot. Malli järjestää luokiteltavat objektit luokitteluperusteen mukaiseen suuruusjärjestykseen ja aloittaa luokittelun siitä objektista, jolla on suurin luokitteluperuste.

Aloitin vertailemalla eri luokittelutapoja ja yritin valita niistä sellaisen, joka parhaiten luokittelee tuotteet. Mielestäni toimivin ratkaisu on se, että tuotteet luokitellaan eron absoluuttisen arvon mukaan. Tässä ratkaisussa voidaan rajat asettaa niin, että ne valitaan hyväksyttävän virheen mukaan ja arvoja verrataan niihin, eikä toisiinsa. Tällöin luokkiin jaotellaan kaikki ne tuotteet, jotka täyttävät tietyn ehdon, eikä muiden tuotteiden onnistuminen vaikuta niihin. Jos kaikki tuotteet on ennustettu hyvin, silloin huonoimmat luokat jäävät tyhjiksi.

Luokitteluperusteen kumulatiivisessa osuudessa luokittelu suoritetaan sillä tavalla, että mitä tasaisemmin tuotteiden ennustaminen on onnistunut, sitä enemmän tuotteita on huonoimmassa luokassa. Tämä johtuu siitä, että yksittäiset suuret erot joidenkin tuotteiden kohdalla kerryttävät nopeammin kumulatiivista osuutta yhteenlasketuista eroista, kuin suurimmat arvot, silloin kun ne poikkeavat vain vähän parhaista tuloksista. Lisäksi suhdelukujen yhteenlasku ei tuota käyttökelpoista arvoa.

Tuotteiden määrään perustuvat luokittelutavat määrittävät luokkien kokoa ja niissä tietty määrä tai osuus tuotteista määritetään kuuluvaksi kuhunkin luokkaan. Näissä siis muiden ennusteiden onnistuminen vaikuttaa tuotteen luokkaan.

Kyseisellä prosessilla ABC-luokitukseen tuodaan datajoukko, jossa on 400 eri tuotetta. Suhdeluvut sijoittuvat välille 0—30,08 ja niiden keskiarvo on 0,60. Arvojen keskihajonta on 2,19. Niillä tuotteilla, joilla suhdeluku on yli 1, ero on keskimäärin suurempi kuin toteuma.

Aloitin testaamisen niin, että tuotteet jaettiin kolmeen eri luokkaan A, B ja C. Lähdin etsimään rajoja, jotka jaottelisivat tuotteet niin, että luokka A olisi suurin ja siinä on tuotteet, joiden ennuste on onnistunut hyvin ja B-luokka olisi toiseksi

suurin ja siinä olisi tuotteet, jotka on ennustettu kohtalaisesti. C-luokkaan halusin luokitella tuotteet, joiden ennuste on onnistunut huonosti. Parin luokittelukerran jälkeen lisäsin vielä luokan D, joka jaotteli luokan C kahteen pienempään luokkaan. ABC-luokitukseen vietävää dataa tutkimalla oli helppo löytää sellaiset raja-arvot, jotka toivat haluamani lopputuloksen. Taulukossa 1 on kuvattuna luokkien raja-arvot ja luokkien koostumukset.

Taulukko 1. Tuotekohtaisen ennusteen tarkkuuden luokat.

Luokka	Rajat	Objektien määrä	Objektien %-osuus	KA	Keski-hajonta	Min. arvo	Max. arvo
A	0-0,3	208	52,00	0,16	0,07	0,00	0,30
B	0,3-0,6	93	23,25	0,44	0,08	0,30	0,59
C	0,6-1,0	77	19,25	0,75	0,08	0,60	0,94
D	yli 1,0	22	5,50	4,90	8,15	1,01	30,08
Yht.		400	100,00	0,60	2,19	0,00	30,08

Seuraavaksi testasin arvioitavan luvun laskutapaa, joka käytti toteuman mukaan painotettua keskiarvoa. Prosessin alku on samanlainen kuin edellisessäkin testissä. Toteuman ja ennusteen välisen eron itseisarvon laskemisen jälkeen, prosessi jatkuu erilaisena. Seuraavaksi ero kerrotaan toteumalla. Eri viikkojen nämä arvot ja toteumat summataan versiokohtaisiksi arvoiksi. Painotettu keskiarvo lasketaan näiden summien osamääränä. Painotettu keskiarvo suhteutetaan vielä jakamalla se toteumien summalla, jotta arvot ovat vertailukelpoiset eri tuotteiden välillä.

Arviointiin käytettävän tunnusluvun laskemisessa on seuraavat vaiheet:

- $Ero = |Toteuma - Ennuste|$, joka lasketaan erikseen kullekin tuote—viikko—versio-yhdistelmälle
- $Kerroin = Ero * Toteuma$
- $SUM(Toteuma) = Toteuma_{vkoN} + Toteuma_{vkoM} + \dots$
- $SUM(Kerroin) = Kerroin_{vkoN} + Kerroin_{vkoM} + \dots$

- Painotettu keskiarvo_{Versio} = SUM (Kerroin) / SUM (Toteuma) / SUM (Toteuma)
- Painotettu keskiarvo_{Tuote} = (Painotettu keskiarvo_{VersioX} + Painotettu keskiarvo_{VersioY} + ...) / COUNT (Versiot)

Koska suhteuttamiseen käytetään analysoitavan ajanjakson kokonaistoteumaa, ajanjakson täytyy olla aina samanmittainen. Rajasin käsiteltävän ajanjakson viikoille 3—11, joilla kaikissa versioissa on ennuste. Rajausta on tehty suodattamalla heti tietolähteiden yhdistämisen jälkeen. Suhdeluvusta tehdään prosenttiarvo kertomalla se sadalla. Lopuksi arvoista lasketaan tuotekohtainen keskiarvo.

ABC-luokituksen luokittelutapa voidaan valita luokitteluperusteen absoluuttiseksi arvoksi samoilla perusteluilla kuin aikaisemmassa prosessissa.

Tässä prosessissa ABC-luokitukseen syötetään dataa, jossa on 392 tuotetta. Tuotteiden eri määrä verrattuna edelliseen johtuu siitä, että käytettävän datan ajanjakso on rajattu eri tavalla sen varmistamiseksi, että ajanjakso on kaikissa versioissa samanmittainen. Arvot sijoittuvat välille 0—3008,16 ja niiden keskiarvo on 13,84. Arvojoukon keskihajonta on 152,02.

Koska tämä on vastaava analyysi, kuin edellinen, ainoastaan arvot, joiden perusteella analyysi tehdään, ovat erilaiset, halusin määritellä raja-arvot samalla tavoin kuin aikaisemmassakin testissä.

Taulukko 2. Painotetulla keskiarvolla lasketun tuotekohtaisen ennusteen tarkkuuden luokat.

Luokka	Rajat	Objektien määrä	Objektien %-osuus	KA	Keskihajonta	Min. arvo	Max. arvo
A	0-5	241	61,48	2,07	1,19	0,00	4,90
B	5-10	103	26,28	7,71	1,41	5,01	9,92
C	10-20	29	7,40	13,66	2,99	10,21	19,57
D	yli 20	19	4,85	196,66	664,37	20,40	3008,16
Yht.		392	100,00	13,84	152,02	0,00	3008,16

Valitsin raja-arvot taas tutkimalla ABC-luokitukseen syötettävää dataa. ABC-luokituksen raja-arvot sekä luokkien koostumukset on esitetty taulukossa 2.

Toimeksiantajani oli kiinnostuneempi jälkimmäisestä ratkaisuvaihtoehdosta, joten lähdin kehittämään sitä eteenpäin. Kehitin myös vastaavan ratkaisun asiakaskohtaiselle analyysille.

Asiakaskohtaisessa analyysissä prosessi oli muuten samanlainen kuin tuotteella, mutta asiakas-kenttä korvasi tuotteen kaikissa kohdissa. Lisäksi otin tässä prosessissa huomioon sen, että sama asiakas voi ostaa eri yksiköissä olevia tuotteita, joten toin kuutioista myös yksiköt ja käytin niitä yhtenä avainkenttänä yhdistämisvaiheessa. Eri yksiköiden arvot yhdistetään samalla, kun lasketaan suhteellisista arvoista asiakaskohtainen keskiarvo.

Tässä prosessissa ABC-luokitukseen tuotava data on kuitenkin erilaista kuin tuotteella. Asiakkaita on ainoastaan 35. Datan koostumus on kuvattu taulukon 3 viimeisellä rivillä.

Taulukko 3. Painotetulla keskiarvolla lasketun asiakaskohtaisen ennusteen luokat.

Luokka	Rajat	Asiakkaiden määrä	Asiakkaiden %-osuus	KA	Keskiahajonta	Min. arvo	Max. arvo
A	0-10	13	37,14	4,51	2,23	1,08	8,24
B	10-100	18	51,43	21,89	19,66	10,41	86,58
C	yli 100	4	11,43	202,98	20,80	171,97	230,36
Yht.		35	100,00	36,13	62,51	1,08	230,36

Koska datan koostumus on erilaista kuin tuotteella, ABC-luokituksenkin asetukset ovat vähän erilaiset. Luokittelutavaksi valitsin edelleen luokitteluperusteen absoluuttiset arvot samoilla perusteilla kuin aiemminkin. Päätin kuitenkin luokitella asiakkaat ainoastaan kolmeen luokkaan, koska niitä on niin paljon vähemmän. Ehkä kaksikin luokkaa olisi riittänyt, mutta päädyin kuitenkin kolmeen,

koska arvot jaottuivat niin, että luokkien raja-arvot olivat melko selkeät. Luokat on kuvattu taulukossa 3.

6.6 Testaus

Kun olin löytänyt sopivat mallit, joilla ongelma voitiin ratkaista ja minulla oli prosessit valmiina datan käsittelyyn ja mallien käyttämiseen, halusin vielä tarkastaa ratkaisujen toimivuuden. Sain koostumukseltaan alkuperäistä testidataa vastaavaa dataa eri ajankohdalta. Uudessa datajoukossa oli toteumaa vuoden 2013 viikoilta 23 – 41 ja ennusteversiot viikoilta 23 – 28.

Rajasin kummassakin prosessissa mukaan ainoastaan viikot 33—41, jotta ajanjakso on yhtä pitkä kuin alkuperäisessä tutkimuksessa. Johtuen siitä, että painotettu keskiarvo suhteutetaan kokonaistoteumalla, arvot ovat vertailukelpoiset ainoastaan samanmittaisilla ajanjaksoilla, kuin ne on asetettu eli tässä tapauksessa yhdeksän viikkoa. Tämän vuoksi muokkasin analyysiprosessien suodattimia ja laskin tulokset uudelleen.

Taulukko 4. Testausvaiheessa painotetulla keskiarvolla lasketun tuotekohtaisen ennusteen tarkkuuden luokat.

Luokka	Objektien määrä	Objektien %-osuus	KA	Keskihajonta	Min. arvo	Max. arvo
A	311	69,11	2,19	1,28	0,00	4,98
B	90	20,00	6,50	1,13	5,01	9,87
C	29	6,44	14,06	2,58	10,20	18,93
D	10	4,44	97,85	154,42	20,38	714,83
Yht.	450	100,00	8,07	38,04	0,00	714,83

Taulukossa 4 on kuvattuna vastaavat tiedot tämän uudemman ajon jälkeen, jotka on kuvattu taulukossa 2 alkuperäiselle datalle. Niitä vertailemalla voidaan päätellä, että jälkimmäisen analyysin tuotteet on ennustettu jonkin verran paremmin, kuin alkuperäisessä analyysissä. Luokkien raja-arvot kuitenkin mieles-

täni olivat sopivat, enkä muuttanut niitä tämän testauksen perusteella. D-luokan ja kokonaistulosten keskiarvoon ja keskihajontaan vaikutti etenkin se, että arvojoukon suurin arvo oli yli 2000 prosenttiyksikköä pienempi jälkimmäisessä analyysissä.

Taulukko 5. Testausvaiheessa painotetulla keskiarvolla lasketun asiakaskohtaisen ennusteen luokat.

Luokka	Asiakkaiden määrä	Asiakkaiden %-osuus	KA	Keskihajonta	Min. arvo	Max. arvo
A	18	52,94	4,02	2,29	0,44	7,78
B	8	23,53	46,66	31,38	11,02	87,57
C	8	23,53	1418,28	2884,98	149,06	9041,22
Yht.	34	100,00	346,82	1520,58	0,44	9041,22

Taulukossa 5 on vastaavasti esitetty tuoreemmalla datajoukolla lasketut tulokset asiakaskohtaiselle painotetulle keskiarvolle. Verrattuna taulukkoon 3 taulukossa 5 arvojen hajonta on paljon suurempi, johtuen yksittäisestä asiakkaasta, jonka kohdalla virhe on yli 9000 %. Yleisesti ottaen myös asiakkaat on onnistuttu ennustamaan paremmin. Tähän datajoukkoon määrittämäni rajat sopivat jopa paremmin kuin alkuperäiseen, joten en niitäkään enää muokannut.

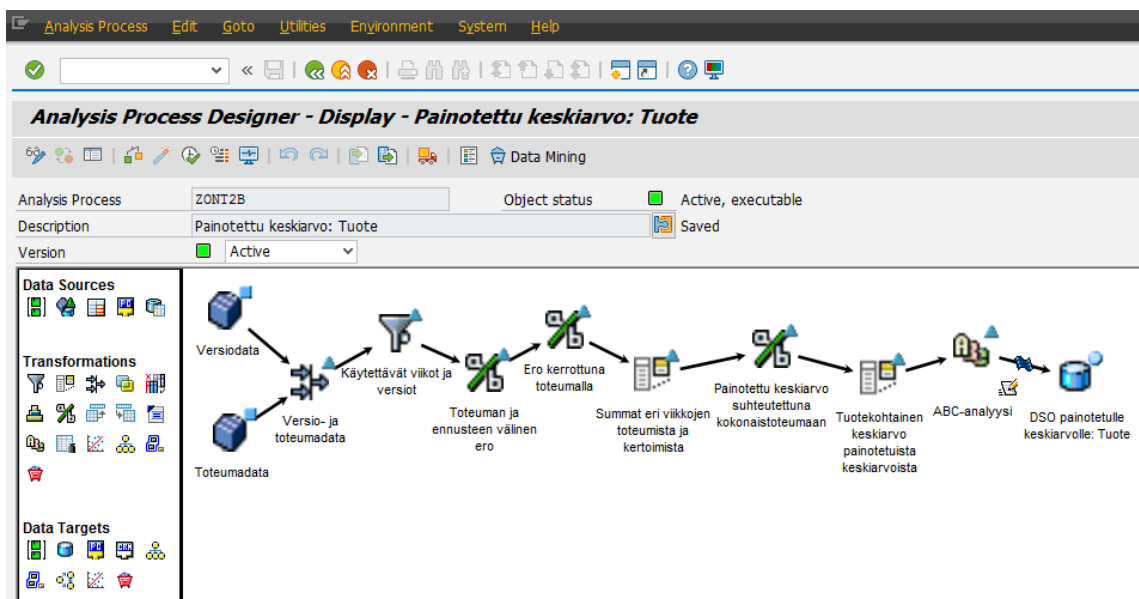
Jälkimmäisissä analyyseissä osa tuotteista ja asiakkaista oli eri luokassa kuin aikaisemmissa analyyseissä. Lähes kaikilla objekteilla vertailuarvot muuttuivat jonkin verran analyysien välillä.

Tämän testin perusteella prosessit toimivat hyvin tämän tyyppiseen analyysiin. Analyysiprosesseille tai tiedonlouhintamalleille ei tarvitse enää tehdä muutoksia, joten voin viimeistellä prosessit.

6.7 Valmiit prosessit

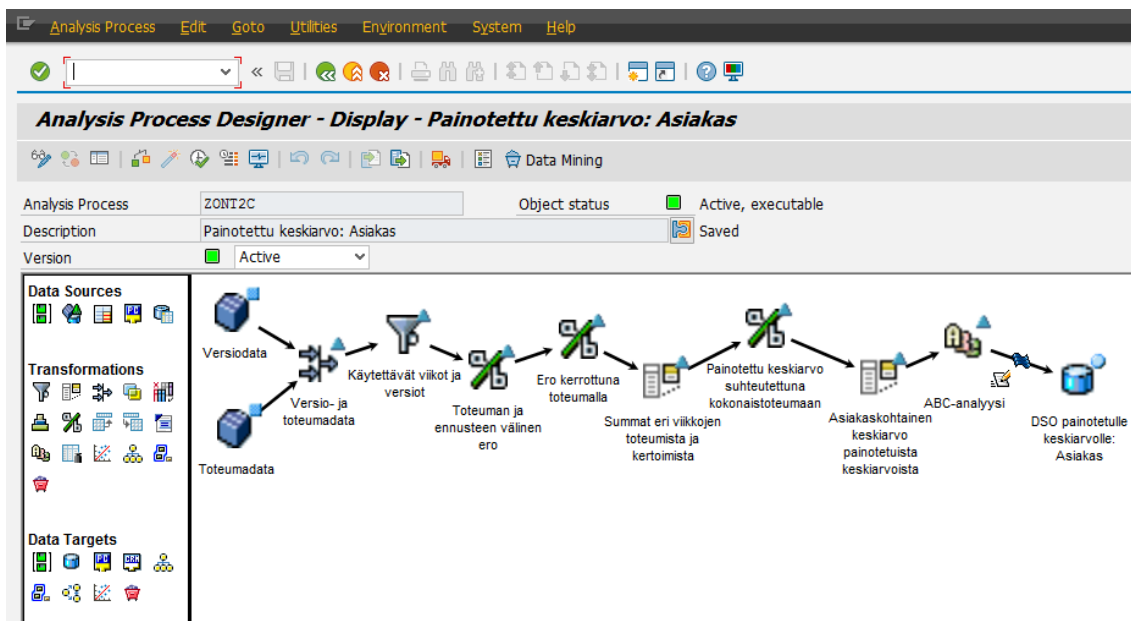
Lopulliset prosessit ovat enimmäkseen samanlaiset kuin kuvasin luvussa 6.5. Ainoastaan tietokohde on muuttunut. Vaikka CSV-tiedostot olivat hyviä tietokoh- teita testausvaiheessa, eivät ne lopullisen tuloksen säilömiseen ole kovin hyviä. CSV-tiedostoissa olevaa dataa ei pysty käsittelemään yhtä monipuolisesti, kuin järjestelmässä olevaa dataa. Tiedostojen saatavuutta ei pysty myöskään var- mistamaan niin tehokkaasti, kuin järjestelmässä olevan datan.

Lopullisissa prosesseissa on tietokohteena DataStore-objekti. Ne ovat tyypil- tään ”Suora kirjoittaminen”, koska analyysiprosessissa datan kirjoittaminen vaa- tii sen. Tuotteen DataStore-objektissa on info-objektit tuote, painotettu keskiarvo ja luokka. Asiakkaan DataStore-objektissa on info-objektit asiakkaalle, painote- tulle keskiarvolle ja luokalle.



Kuva 4. Analyysiprosessi tuotekohtaiselle painotetulle keskiarvolle.

Kuvassa 4 on tuotteen analyysiprosessi. Vasemmassa reunassa on analyysi- prosessisuunnittelijan tietolähteet, muunnokset ja tietokohteet. Analyysiprosessissa näkyy eri solmut, jotka tarvitaan datan käsittelyyn, tiedonlouhintaan datan la- taamiseen DataStore-objektiin. Asiakkaan prosessissa on samat solmut, joiden sisältö joiltain osin vaihtelee. Asiakkaan analyysiprosessi on kuvattu kuvassa 5.



Kuva 5. Analyysiprosessi asiakaskohtaiselle painotetulle keskiarvolle.

Loin kummallekin analyysiprosessille prosessiketjun, joka suorittaa ensin kyseessä olevan analyysiprosessin, poistaa infokuutiossa olevan datan ja sen jälkeen lataa datan DataStore-objektista infokuutioon. Infokuutioissa dataa pysyy hyödyntämään esimerkiksi raportoinnissa Business Explorer Analyzerissa. Infokuutioiden rakenne on samanlainen kuin sitä vastaavan DataStore-objektin. Tuotteella on aktivoituna navigointiattribuutti suunnittelija, joten tuotteille saadaan tarvittaessa myös suunnittelijatieto esim. raportteihin.

Lopuksi tein tulosten esittämistä varten Business Explorer Query Designerilla kyselyt, joilla haettiin kaikki data infokuutioista. Lisäksi loin laskennallisen tunnusluvun lukumäärä, jonka avulla raportissa pystyy laskemaan objektien lukumäärän luokissa. Kyselyn poikkeukset värjäävät eri luokkiin kuuluvien objektien rivit eri väreillä.

PK: Tuote - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Developer Add-Ins PDF-XChange 4

B4

Table				
Tuote	Suunnittelija nav.	Luokka	Painotettu keskiarvo	Lukumäärä
10	005	B	5,38 %	1
11	005	B	5,63 %	1
12	001	B	5,41 %	1
13	009	B	6,49 %	1
14	003	A	1,09 %	1
15	001	D	22,21 %	1
17	008	A	0,00 %	1
24	001	A	4,32 %	1
25	008	B	6,49 %	1
27	004	A	4,37 %	1
28	009	B	5,59 %	1
29	006	A	1,01 %	1
30	006	A	3,56 %	1
31	009	A	2,83 %	1
32	009	B	5,55 %	1
33	010	A	2,34 %	1
34	004	A	1,23 %	1
36	004	A	3,03 %	1
37	009	A	2,72 %	1
38	Not assigned	A	1,76 %	1
39	006	A	1,82 %	1
41	Not assigned	A	3,96 %	1
76	002	A	1,24 %	1
79	009	D	302,01 %	1
106	003	A	3,21 %	1
107	008	D	20,38 %	1
108	008	D	102,56 %	1
109	001	D	69,10 %	1
112	006	A	3,20 %	1
113	010	D	41,11 %	1
114	006	B	5,48 %	1
115	007	A	3,24 %	1
118	002	A	3,54 %	1
119	002	D	714,83 %	1
128	005	A	4,58 %	1
129	001	C	15,97 %	1
132	004	A	0,54 %	1
133	001	B	5,22 %	1

Luokka

☐ #

☐ A

☐ B

☐ C

☐ D

Submit Clear

Suunnittelija nav.

☒ All

☐ #

☐ 001

☐ 002

☐ 003

☐ 004

☐ 005

☐ 006

☐ 007

☐ 008

☐ 009

☐ 010

Table

Luokka

Lukumäärä

A

311

B

90

C

29

D

20

Ready

Kuva 6. Tuotekohtaisen painotetun keskiarvon raportti.

Business Explorer Analyzerilla tein kuvassa 6 esiintyvän raportin. Taulukossa on kuvattu osa tuotekohtaisista tuloksista. Keskellä on muutamia työkaluja, joilla voi valita mitä tietoja halutaan tarkastella ja poistaa eri luokkien väritykset. Oikeanpuolimmainen taulukko käyttää pohjana samaa kyselyä, mutta siitä on poistettu tasot tuote ja suunnittelija, ja tulokseksi on saatu tuotteiden lukumäärä kussakin luokassa. Raportin tulokset on laskettu tuoreemmalla datajoukolla, eli testausvaiheen tuloksilla. Asiakaskohtaiselle painotetulle keskiarvolle tein vastaavan näköisen raportin ilman suunnittelija-kenttää.

7 YHTEENVETO

Tutkimusaiheena tiedonlouhinta on laaja kokonaisuus, joten rajasin myös teoriaosuuden tarkoituksella keskittymään SAP-järjestelmään ja sen tarjoamiin metodeihin. Onnistuin mielestäni melko hyvin kokoamaan perustiedot tiedonlouhinnasta, prosessista ja SAP:n metodeista työn teoriaosuuteen.

Myös myynninsuunnittelu on itsessään laaja aihe, josta olisi helposti saanut opinnäytetyölle aiheen. Koska tässä työssä oli tarkoitus keskittyä tiedonlouhintaan ja myynninsuunnittelua käsittelevän luvun tarkoitus on avata lukijalle käytännön osuuden päämääriä, halusin pitää luvun melko lyhyenä. Halusin kuvata myynninsuunnittelun pintapuolisesti niin, että lukija ymmärtää, miksi myynninsuunnittelun onnistumista on tarpeen arvioida.

Käytännön osuuden tavoitteena oli löytää malli, jonka avulla voidaan analysoida tuotteen ja asiakkaan myyminenennusteen onnistumista. Löysin kaksi mahdollista vaihtoehtoa, joiden avulla kyseinen analyysi voidaan suorittaa. Lisäksi pystyin perustelemaan, miksi muut mallit eivät sovellu tähän tarkoitukseen.

Opinnäytetyöprosessissa olisi mielestäni jonkin verran parannettavan varaa. Kriittikkiä antaisin itselleni ainakin työn aikatauluttamisesta, koska en saanut työtä suoritettua siinä aikataulussa, jota olin suunnitellut. Myös työn rajausta olisi kannattanut miettiä tarkemmin etukäteen, koska käytännön osuus oli melko laaja ja rajasin jälkikäteen suuren osan siitä raportin ulkopuolelle.

Tässä työssä oli tavoitteena tutkia ainoastaan tiedonlouhinnan hyödyntämistä analyysissa, joka määrittää onko tietyn objektin myyminenennuste onnistunut hyvin vai huonosti. Tiedonlouhinnasta kuitenkin löytyisi myös monipuolisia työkaluja esimerkiksi hyvään tai huonoon tulokseen johtaneiden syiden löytämiseen.

Hyödyt toimeksiantajalle eivät olleet ihan sitä, mitä he olivat odottaneet, koska useimmat mallit eivät soveltuneet ennusteversioiden analysointiin halutulla tavalla ja dataa joutui prosesseissa käsittelemään melko paljon ennen varsinaista tiedonlouhinta. Analyttisistä malleista ABC-luokitus ja painotetut pisteytystau-

lukot osoittautuivat kuitenkin hyödyllisiksi versiodatan analysoinnissa. Lisäksi analyysiprosessisuunnittelijan käytöstä saadut tiedot osoittautuivat hyödyllisiksi ja sitä voidaan mahdollisesti hyödyntää myös muunlaisissa tehtävissä kuin tiedonlouhinnassa.

LÄHTEET

Dickersbach, J. T, 2009. Supply Chain Management with SAP APO : structures, modeling approaches and implementation of SAP SCM 2008. 3rd Edition. Dordrecht : Springer.

Han, J. & Kamber, M. 2006. Data Mining : Concepts and Techniques. 2nd Edition. Amsterdam ; London : Elsevier.

Linoff, G. S. & Berry, M. J. A. 2011. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 3rd Edition. Indianapolis, Ind. : Wiley Pub., Inc.

McDonald, K.; Wilmsmeier, A.; Dixon D. C. & Inmon W. H. 2006. Mastering the SAP Business Information Warehouse. Second Edition. Indianapolis, IN: Wiley Pub.

Prabhu, S. & Venkatesan, N. 2007. Data Mining and Warehousing. New Delhi : New Age International (P) Ltd., Publishers.

SAP Help Portal 2013a. ABC Classification. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/51/9f883c119d5d71e10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013b. Association Analysis. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/da/ec293b31de281de10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013c. Classification Rules. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/af/f2853c80c7554ae10000000a11405a/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013d. Clustering. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/c5/e5293b5fdebb1ce10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013e. Data Mining. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/4a/eb293b31de281de10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013f. Decision Trees. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/72/e5293b5fdebb1ce10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013g. Demand Planning. Viitattu 24.11.2013
http://help.sap.com/saphelp_scm70/helpdata/EN/8a/9d6937089c2556e100000009b38f889/frameset.htm.

SAP Help Portal 2013h. Demand Planning Process. Viitattu 24.11.2013
http://help.sap.com/saphelp_scm70/helpdata/EN/99/809d9be771456d83b8e24f9ef140e8/frameset.htm.

SAP Help Portal 2013i. Regression Analysis. Viitattu 24.11..2013
http://help.sap.com/saphelp_nw73/helpdata/en/9d/221640b35ec442e10000000a1550b0/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013j. Scoring. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/a1/e5293b5fdebb1ce10000000a114084/content.htm?frameset=/en/4f/bbfd3f0521c842e10000000a1550b0/frameset.htm.

SAP Help Portal 2013k. Special Settings for Weighted Score Tables. Viitattu 24.11.2013
http://help.sap.com/saphelp_nw73/helpdata/en/4f/bbfd3f0521c842e10000000a1550b0/content.htm.

Snapp, S. 2010. Discover SAP SCM. Boston : Galileo Press Inc.

Tufféry, S. 2011. Data Mining and Statistics for Decision Making. Translated by Rod Riesco. Chichester, West Sussex ; Hoboken, NJ. : Wiley.